8Л

УДК 004.896

О.Н. Лозович[1], А.Л. Максимов[2]
[1]Государственный университет информатики и искусственного интеллекта,
г. Донецк, Украина
[2]Одесский национальный университет им. И.И. Мечникова, Украина
simmetriada@yandex.ru, teerle@te.net.ua

Применение нейронных сетей на основе ПЛИС для решения задач надежности коммуникационных информационных систем

В статье рассматривается проблема построения нейронных сетей для решения задачи надежности информационных систем. Предлагается построение рекуррентных нейронных сетей и фильтров Калмана на основе программируемых интегральных схем.

Введение

Большинство неисправностей, возникающих в информационных системах, малоформализованы и трудно поддаются предсказанию, основанной на традиционных математических методах. Каждый компонент информационной системы может иметь свой коэффициент надежности, что ведет к большому числу входной информации и громоздким решениям [1-4].

Применяя парадигму нейронных сетей, можно создавать параллельные и распределенные структуры с высокой скоростью вычислений и большей эффективностью для реализации задач контроля и обеспечения надежности информационных систем. Для реализации распараллеливаемых моделей нейронных сетей возможно использовать программируемые логические интегральные схемы (ПЛИС).

В статье предложен вариант реализации нейронных сетей на программируемых логических элементах. Большую часть элементов типичных сетей можно построить из реализуемых в ПЛИС арифметических функций (сложение, вычитание, умножение, деление, вычисление квадратного корня), компараторы. Это, в частности, позволяет построить внутреннюю реализацию достаточно гладких аппроксимаций сигмоидальных нелинейных функций. Кроме того, существует возможность использования современных интерфейсных решений в ПЛИС, связанных с управлением и программированием формы и скорости входных и выходных сигналов.

Большое количество вентилей ПЛИС позволяет реализовать достаточно много физически параллельно работающих нейронов, обмен данными между которыми может осуществляется внутри ПЛИС. Каналы связи между логическими элементами и внутренней памятью ПЛИС обладают хорошими скоростными характеристиками. Высокая тактовая частота работы ПЛИС (в современных схемах – до гигагерцевых частот), параллельность функционирования логики с хорошими внутренними и внешними каналами передачи данных способствует высокой скорости вычислений для множества параллельно функционирующих нейронов (а точнее, их реализаций).

Учитывая особенности ПЛИС, в статье рассматриваются карты самоорганизации и рекуррентные сети и фильтры Калмана для реализации двухуровневой системы, решающей задачи отслеживания и детектирования опасных состояний и повышения

надежности в коммуникационных информационных системах. На верхнем уровне предлагается использование карт самоорганизации Кохонена, которые распознают тип поведения системы на глобальных периодических интервалах времени. На нижнем уровне предлагается использование фильтров Калмана для предсказания и оценки поведения системы на небольших интервалах времени. Также предложено ввести оценку параметров d, определяющих отклонение работы коммуникационной информационной системы от эталонной или типовой модели.

Цель данной работы состоит в разработке методов и средств повышения надежности коммуникационных информационных систем посредством построения нейронных сетей на базе ПЛИС.

Объектом исследования является программно-аппаратная структура, определяемая используемыми моделями нейронных сетей и ПЛИС.

Предметом исследования является оценка реализуемости двухуровневой нейронной модели посредством параллельно-конвейерных вычислений в ПЛИС заданных типов.

В данной работе исследуются также математические модели и алгоритмы обучения нейронных сетей, представленных в базисе ПЛИС.

Вычислительные структуры в ПЛИС

Программируемые логические интегральные схемы состоят из логических блоков и коммутирующих связей – программируемых матриц соединений. Логические блоки таких ПЛИС состоят из одного или нескольких относительно простых логических элементов, в их основе лежит таблица перекодировки, программируемый мультиплексор, триггер, цепи управления. Таких простых элементов может быть до 1 миллиона, число логических элементов достигает нескольких десятков тысяч. За счет большого числа логических элементов предоставляется возможность реконфигурировать модули памяти, что делает ПЛИС средством реализации алгоритмов цифровой обработки сигналов, основными операциями которых являются перемножение, умножение на константу, суммирование, задержка сигнала. Характерными для архитектур FPGA являются элементы ввода – вывода, позволяющие реализовать двунаправленный ввод – вывод, третье состояние.

Логические элементы объединяются в группы – логические блоки. Внутри логических блоков логические элементы соединяются посредством локальной программируемой матрицы соединений, позволяющей соединять любой логический элемент с любым другим. Логические блоки связаны между собой и с элементами ввода – вывода посредством глобальной программируемой матрицы соединений. Локальная и глобальная матрицы соединений имеют непрерывную структуру, для каждого соединения выделяется непрерывный канал.

ПЛИС фирмы «Альтера» выпускаются с возможностью программирования в системе непосредственно на плате.

Моделирование вычислительных структур устройств, функционирующих в базисе ПЛИС

За счет простоты реализации обратных связей в ПЛИС рекуррентные нейронные сети являются одним из важнейших типов нейронных сетей, которые возможно реализовывать на таких интегральных схемах.

Рекуррентные нейронные сети возможно использовать для решения задач аппроксимации временных последовательностей, классификации. Рекуррентные сети

являются динамическими, так как в силу обратных связей в них модифицируются входы нейронов, что приводит к изменению состояния сети. Поведение рекуррентных сетей описывается дифференциальными или разностными уравнениями, как правило, первого порядка. Это гораздо расширяет области применения нейросетей и способы их обучения. Сеть организована так, что каждый нейрон получает входную информацию от других нейронов, возможно, и от самого себя, и от окружающей среды.

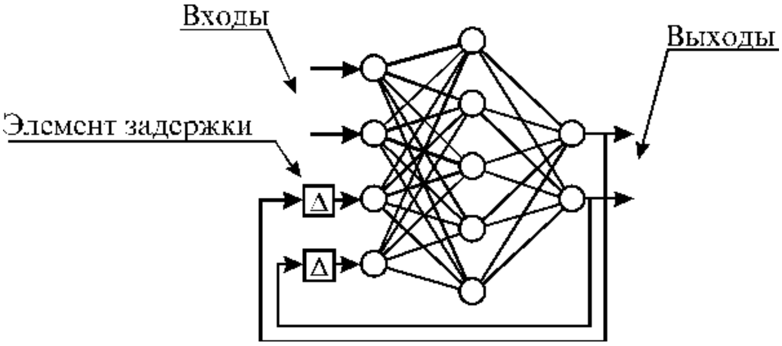


Рисунок 1 – Рекуррентная сеть с одним скрытым слоем

Учитывая возможности сред разработки для современных ПЛИС (в частности, в них возможна автоматическая реализация умножителей, сумматоров, делителей разных типов, вычисление квадратного корня и других мега-функций), появляется возможность реализации нейронных сетей типа самоорганизующихся карт Кохонена.

В картах Кохонена весовые коэффициенты перевычисляются по следующей формуле:

$$w_j(n+1) = w_j(n) + \eta(n) h_{i,j(x)}(n)(x - w_j(n)), \quad (1)$$

где $w_j(n)$ – вес j-го нейрона на n шаге,

$\eta(n)$ – параметр скорости обучения,

$h_{i,j(x)}(n)$ – функция окрестности с центром в победившем нейроне,

$j(x)$ – вектор наблюдения на n шаге.

Поиск максимального подобия вычисляется поиском наиболее подходящего нейрона $j(x)$ на шаге n, используя критерий минимума Евклидова расстояния:

$$f(x) = \arg\min_j \left\| x - w_j \right\|, j = 1,2,...,n. \quad (2)$$

В пространстве векторов выбирается набор векторов, которые близки к объектам.

Признаки надежности систем зависят от времени. Для времязависимых входных векторов удобно считать пространство, содержащее наблюдаемые векторы x, функциональным пространством или отображением, у которого область определения является временем, а область значений может быть гетерогенным пространством, состоящим из двух частей – наблюдаемой статистической части и оценкой текущего состояния системы.

Для распознавания (классификации) состояний коммуникационной информационной системы необходимо адекватное представление параметров времени, статистики, а также дополнительного параметра, представляющего себе внешнюю оценочную характеристику состояния коммуникационной информационной системы. Эта характеристика может быть задана экспертом или получаться автоматически на основании обработки системных журналов.

Для параметра времени следует определить циклы, которые коррелируют с состоянием системы внутри временных интервалов и определяются задачами функционирования системы. Обычно существуют циклы задаваемые внутри суток, по неделям, по месяцам, годовая цикличность. Необходимость учета наблюдаемой цикличности в поведении системы привела к следующему решению: мы предлагаем одномерное временное пространство погрузить в многомерный тор. При представлении многомерного тора в евклидовом пространстве время можно представить в виде кривой на поверхности многомерного тора, погруженного в евклидово пространство достаточной размерности. Многомерный тор описывается как топологическое произведение соответствующих циклических подпространств. В нашей работе мы выделяем шесть таких подпространств:

$$T = S^1 \times S^2 \times S^3 \times S^4 \times S^5 \times S^6, \qquad (3)$$

где T – топологическое пространство для представления времени, S^1 – временной интервал порядка часа, S^2 – суточный интервал, S^3 – недельный интервал, S^4 – месяц, S^5 – квартал, S^6 –год.

Каждое из этих подпространств (на упрощенном рис. 2 они представлены выделенными кольцами) – одномерное замкнутое пространство. Это означает, что достаточно легко реализовать вычисления многомерной близости для параметров времени. Вычислительные операции в нейронной сети, которые определяют близость расстояния, должны быть модифицированы в соответствии с представленной топологической моделью, которая определяет отношение близости во временном пространстве с учетом естественных периодических факторов.

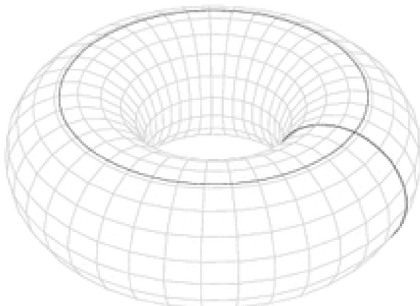


Рисунок 2 – Вложение двухмерного тороидального топологического временного пространства в трехмерное пространство

Изменение системы, выходящее за рамки адаптивных отслеживающих нейронных сетей для заданного интервала времени, примем за неисправность, а также примем за неисправность уход параметров от эталонной модели за пределы допустимых отклонений. Цикличность на коротких временных интервалах (меньших часа) зависит от работы системы и может меняться.

Статистика привязана к данной топологической модели времени, т.к. она также зависит от времени. Параметры статистики представляются в виде вектора в обычном евклидовом пространстве. На выходе есть предварительная оценка качества функционирования системы. Задача нейронной сети, реализующей карты Кохонена, состоит в том, чтобы минимизировать получаемую информацию и построить наборы кластеров, которые отражают типичную работу системы в разные моменты времени.

Далее необходимо адаптировать формулу (2) к постановке нашей задачи. Тогда получим:

$$f(x) = \arg\min_j \min(p_t(x, w_j) + p_d), \qquad (4)$$

где элементарное состояние w_j – это вектор, описывающий состояние системы в заданный момент времени;

d – внешняя («экспертная») оценка, которая также является компонентом вектора w_j;

ρ_t – составляющая часть метрики по времени;

ρ_s – составляющая часть метрики по статистическим характеристикам;

ρ_d – вклад внешней оценки в расстояние.

Для оценки расстояний можно использовать евклидовы нормы или модифицированные нормы с учетом топологии пространства.

Таким образом, появляется возможность корректировать веса в формуле (2) с учетом временной зависимости данных и типичных задающих временных интервалов.

Для отслеживания поведения системы на коротких временных интервалах в ПЛИС предпочтительнее использовать рекуррентные фильтры. Одним из наиболее хорошо зарекомендовавших себя рекурсивных нейронных сетей является фильтр Калмана.

Фильтр Калмана должен отслеживать неисправности двух вариантов: быстрые временные изменения, которые нарушают свойства линейной предсказуемости, и изменения, которые уходят от заданной адаптивной модели, определяемой картой Кохонена.

Особенностью фильтра Калмана является его хорошая обучающая адаптивная способность.

Для предсказания поведения системы, целесообразно использовать фильтры Калмана, которые характеризуются:

– формулировкой в терминах концепции пространства состояний и эффективным использованием информации, содержащейся во входных данных;

– рекурсивной оценкой состояний, что означает вычисление каждой скорректированной оценки на основе предыдущей оценки и допустимых в текущий момент данных.

Рассмотрим один из самых простых вариантов использования фильтра Калмана для отслеживания текущего состояния системы.

Уравнение, процесса, описывающего предсказуемую модель изменения текущего состояния на заданном интервале:

$$w(n+1) = Fw(n) + v(n). \tag{5}$$

Уравнение описывающее процесс измерения этого состояния:

$$d(n) = c(n)w(n), \tag{6}$$

где $w(n)$ – вектор состояния системы,

$d(n)$ – вектор наблюдения,

$c(n)$ – матрица измерений,

$v(n)$ – шум или флуктуации в поведении системы,

F – матрица изменений состояния системы на каждом шаге (в одномерном случае – это скалярный параметр, определяющий тип локальной эволюции измеряемой величины).

Точное состояние $w(n)$ системы может меняться в определенных пределах, задающих границы нормального функционирования. Фильтр Калмана хорошо отслеживает локальную эволюцию скрытых параметров и позволяет своевременно обнаруживать статические и динамические отклонения, правда, в рамках линейной модели отслеживания.

Описание фильтра Калмана возможно с помощью понятия инновации. Более точно процесс инновации, связанный с вектором наблюдений $d(n)$, определяется следующим образом:

$$a(n) = d(n) - \hat{d}(n|n-1), \tag{7}$$

где $\hat{d}(n|n-1)$ – оценка $d(n)$ с минимальной среднеквадратической ошибкой на основе всех прошлых значений вектора наблюдений, начиная с момента времени $n = 1$ и заканчивая моментом $n-1$. Под оценкой с минимальной среднеквадратической ошибкой понимается та оценка, которая минимизирует среднеквадратическую ошибку $d(n)$. Инновационный процесс $a(n)$ можно рассматривать как меру новой информации, содержащейся в $d(n)$, которая была недоступна в прогнозируемой части $\hat{d}(n|n-1)$.

Предлагаемая нейросетевая контролирующая система изображена на рис. 3. На начальном этапе определяется тип функционирования системы с помощью сетей Кохонена, на втором этапе выполняется отслеживание отклонения поведения системы на небольших интервалах.

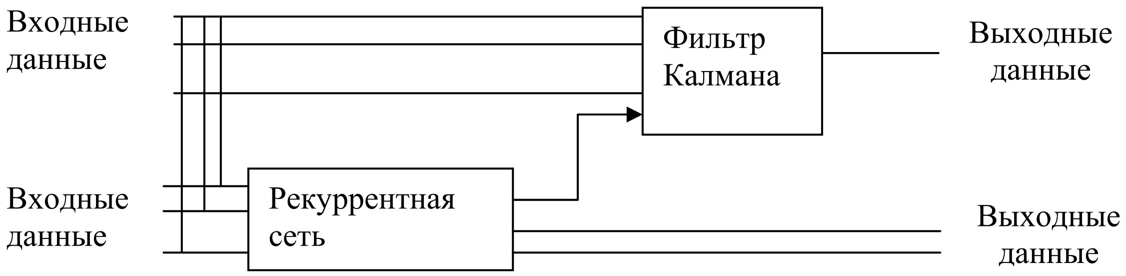


Рисунок 3 – Нейросетевая контролирующая система

Входными сигналами является последовательность изменений параметров как функция времени для каждой компоненты системы.

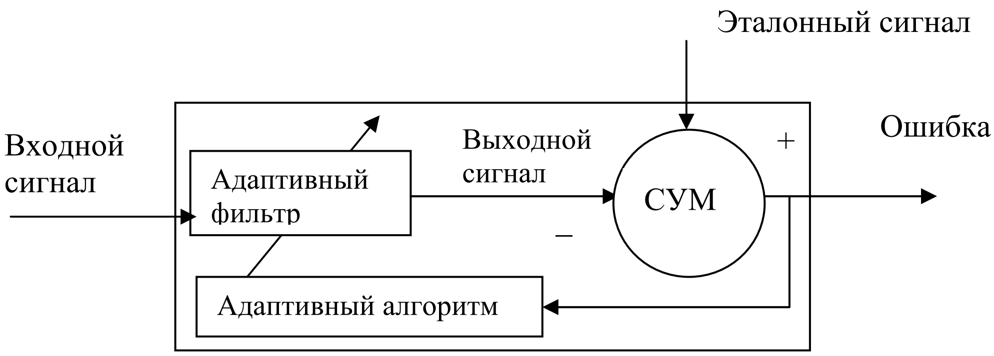


Рисунок 4 – Решение задачи идентификации отклонения поведения от эталона с помощью рекуррентной адаптивной нейронной сети

Критерий нормального функционирования на локальных интервалах определяется локальным или глобальным минимумом на поверхности ошибок рекуррентной сети. Для рекуррентной сети, построенной с W синаптическими весами и p выходными узлами, можно записать уравнение эволюции параметров сети в точки локального или глобального минимумов как:

$$w\,(n+1) = w(n),$$
$$d_o(n) = c(\,w(n),\,u(n),\,v(n)) + v(n), \quad (8)$$

где вектор весов $w(n)$ играет роль состояния (но уже самой сети),

вектор $d_o(n)$ – желаемый отклик сети.

Второй аргумент $u(n)$ и третий аргумент $v(n)$ в нелинейной вектор-функции $c(, ,)$ измерений обозначают соответственно входной вектор и вектор рекуррентной активности узла. В результате уравнение (8) определяет квазиустойчивое состояние синапсов нейронной сети, соответствующее некоторому заданному поведению.

Для большого числа нейросетевых моделей важным является выбор нелинейной функции преобразования. Сигмоидальная функция является одной из важных для нейросетевых систем с хорошими адаптационными свойствами (например, хорошей обучаемости, мягкого отслеживания изменений и т.д.). В стандартной архитектуре ПЛИС отсутствует такая операция среди готовых макрофункций. Однако возможно построить аппроксимацию сигмоидальной функции с помощью кусочно-рациональных функций имеющих разрыв второй производной в точке ноль и допустимую дискретность. Такая аппроксимация должна обеспечивать сходимость методов обучения, если предполагается автономное обучение нейронной сети. Для многослойных и рекуррентных сетей допустима следующая аппроксимация:

$$\varphi(x) = C + \frac{Mkx}{1+|kx|}, \qquad (9)$$

где M – выходной диапазон значений нелинейной функции, область значений;

k – угловой коэффициент;

C – коэффициент, определяющий нулевой уровень (уровень срабатывания, если рассматривать логику компараторов). Для неотрицательных чисел $C = M$.

Исходя из того, что каждая такая функция требует относительно больших затрат ресурсов (логических вентилей и элементов памяти), наиболее выгодными с точки зрения реализации являются рекуррентные схемы нейронных сетей. Для реализации нейросетевых алгоритмов в ПЛИС весовые коэффициенты в матрицах, исходя из ограничений реализации, следует выбирать из целых чисел. Идеальными весовыми коэффициентами являются коэффициенты степени два. Дробные коэффициенты моделируются как рациональные числа. В линейных частях реализации нейромоделей используются числители рациональных чисел. Знаменатель можно вынести в качестве расширенного параметра M в уравнении (9) или учитывать его при нормировании. Это накладывает ограничения на выбор коэффициентов и, в частности, делает желательным процесс предварительного обучения нейронных сетей. Важной задачей является задача разработки робастных алгоритмов автономного обучения для сетей с такими ограничениями.

Выводы

Предложенная модель реализации нейронной сети на базе ПЛИС дает возможность предсказывать сбои в коммуникационных информационных системах на основе фильтра Калмана и карт Кохонана в зависимости от изменений работы системы, учитывая временные интервалы.

Представление входных данных с учетом временных интервалов дает возможность учитывать периодические изменения в работе системы, связанные с внешними факторами.

Применение ПЛИС для реализации нейронных сетей позволяет увеличить скорость обработки сигналов, в отличие от программной реализации.

Литература

1. Саймон Хайкин. Нейронные сети : полный курс / Саймон Хайкин. – [Пер. с англ. ; 2-е изд.]. – М. : Издательский дом «Вильямс», 2006. – 1104 с.
2. Медведев А.И. Теория надежности сложных систем / А.И. Медведев, В.А. Каштанов. – Физматлит, 2010. – 608 с.
3. Комарцова Л.Г. Нейрокомпьютеры : учебное пособие для вузов / Л.Г. Комарцова, А.Г. Максимов. – [3-е изд.]. – М. : МГТУ им. Н.Э. Баумана, 2004. – 400 с.
4. Уоссермен Ф. Нейрокомпьютерная техника. / Уоссермен Ф. – М.: Мир, 1992.
5. Галушкин А.И. Нейрокомпьютеры : учебное пособие для вузов / Галушкин А.И. – М. : ИПРЖР, 2000. – 528 с.

Literatura

1. Neironnye seti: polnyi kurs, 2-e izdanie. Saimon Haikin, Per. s angl. – M. : Izdatel'skii dom Vil'yams, 2006. – 1104 s.
2. Teoriya nadejnosti slojnyh sistem. Medvedev A.I., Kashtanov V.A., Izdatel'stvo: Fizmatlit, 2010 god. – 608 s.
3. Neirokomp'yutery. Komarcova L.G., Maksimov A.G., Uchebnoe posobie dlya vuzov. - vtoroe izdanie - M. Izdaiel'stvo : MGTUim. N.E. Baumana, 2004. – 400 s.
4. Neirokomp'yuternaya tehnika. Uossermen F. M.: Mir, 1992.
5. Neirokomp'yutery. : Uchebnoe posobie dlya vuzov. Galushkin A.I. – M. : IPRJR, 2000. – 528 s.

О.М. Лозович, А.Л. Максимов

Використання нейронних мереж на основі програмованих логічних інтегральних схем для розв’язання задач надійності

У статті розглядається проблема побудови нейронних мереж для вирішення проблем надійності комунікаційних інформаційних систем. Пропонується двохетапний метод побудови системи на основі карт Кохонана та фільтра Калмана.

O.N. Lozovich, A.L. Maksimov

The use of Neuron Network on the Basis of the Programmable logical Integrated Circuits for the Decision of Tasks of Reliability

In the article the problem of construction of neuron networks is examined for the decision of problems of reliability of the of communication informative systems. It is offered two a stage method, constructions of the system on the basis of Kokchen maps and Kalman filter.

Статья поступила в редакцию 31.05.2011.